



Figure S1: Pairwise comparisons in AUC within each evaluation category of GO terms. (a) For each pair of initial submissions X and Y, we test for difference in AUC value for every GO term within a given evaluation category, evaluated using held-out genes. Color bars indicate fraction of pairwise comparisons for which X's AUC is significantly higher (blue), not significantly different (beige) and significantly lower (maroon) than Y's AUC. (b) As (a), except evaluated using the newly-annotated genes. (c) (d) As (a),(b) except that asterisks indicate second-round submissions.